

ICONIC REPRESENTATION OF CONTENT

5

Background Of The Invention

Field of the Invention

- 10 This invention generally relates to computer user interfaces. More specifically, the invention relates to a method and system for improving the searching of computer files via representation of their content as icons.

Prior Art

15

Many functions in modern computers can be very time consuming. From having to wait to turn on the computer, waiting for all programs to load, and then finally having to wait to determine what each file contains. For example, a user is not familiar with a certain computer may want to find a file about car mechanics. That person would probably have to go to a separate directory, such as DOS, Windows Explorer, etc., in order to be able to find a certain file. After reaching the separate directory, the user would have to specify a search on all the files in a drive. After this long search a user would have to go through each file and read about the file and then sooner or later find the file they were looking for.

20

Another method to solve this problem is to conduct a basic search for file names on the computer's operating system. After conducting the search, the user would have to browse through different files and open every file separately and check to determine if the file is the one needed.

25

30

This can be a very long and useless process because of the amount of work needed to open and search through drives and directories. Other searches that compare keywords with text in documents can also be frustrating, especially for beginning users.

- 5 This process is also extremely lengthy and sometimes even pointless when considering the number of files that could show up in one file search. This method is very time consuming because of the fact that before the user finds the file he/she is looking for, they may have to go through opening a large number of files.
- 10 Another method is to use the icons listed throughout the drives and the desktop. By right clicking on an icon, the user can get a basic menu. Typically, one of the options on the menu is “ properties,” which allows the user to view a small number of details about the file.
- 15 This process of trying to find a file is very unlikely to be helpful because it is almost a guess to as what files are the ones needed. The user would have to go through a large amount of files before he/she finds the one needed. This method is also very time consuming because the user would have to go through a number of files and spend a few minutes looking over the details of the files. Another reason why this method is
- 20 not very helpful is because the details listed by the properties function are not very informative about the file's content.

Summary Of The Invention

- 25 An object of this invention is to be able to provide a user, regardless of whether the user is familiar or not familiar with a computer, with an easily accessible method to find programs or any files that he or she needs without taking up too much time or patience.

Another object of the present invention is to provide a versatile system for determining and displaying icons representing files or portions of files containing text such as e-mail, web pages, text documents, word-processor documents, etc.

- 5 A further object of the present invention is to determine the content of the text of a document by examining words in the document, and then choosing the closest icon available and displaying it as the icon representing the text document.

These and other objectives are attained with a method and apparatus for determining
10 and displaying icons representing files containing text, such as e-mail, e-books, web pages, text documents, word-processor documents, etc. In particular, the system determines the content of the text by examining words in the document. For example, if words relating to cars appear several times in the document, then the document's topic probably relates to car. Next, the system searches in a database of icons, which
15 are labeled according to type. For example, the database may contain graphics relating to transportation (cars, planes, trains, etc.) computers (hard disk, monitor, keyboard, etc.), animals (mammal, reptile, amphibian), and many other categories. The system chooses the closest icon available and displays it as the icon representing the text document. (For example, the system may associate the document on cars with
20 a car icon, and the car icon is displayed in appropriate regions of the desktop such as in file listings, desktop shortcuts, menus, task bars, etc.)

One way for the system to select an appropriate icon is by comparing a content word from a document, such as car, with the database of icons, which also contains words
25 associated with each icon. As an example, the database may contain records containing words and names of icon (graphical image) files:

Icon Database:

30 Text Icon file name

car	car.jpg
dog	dog.jpg
keyboard	keyboard.jpg
amphibian	frog.jpg
5 frog	frog.jpg

If the topic word is car, the system searches the text in the icon database for "car." When there is a match, the system reads the icon file name car.jpg and displays the icon. The image car.jpg may include an advertisement.

10

Various methods are available for determining the "content" of a document, or of the sections of a document. Such methods include latent semantic indexing, known to those skilled in the art of content determination, and examination of words in titles and headings, and in the body of a document. For example, if a chapter title in a document contains the word amphibian, the chapter likely is about amphibians, and an amphibian picture (e.g. frog.jpg) may be displayed.

15

As an extension of this basic principle, topic icons may be determined several times in a document. For example, the topic of one paragraph may be cars and another

20

paragraph topic might be trains. These icons may be displayed in the text document so that people can get an idea about content of a document with a quick glance. The icons may also be displayed outside the document so that users can get an idea as to the nature and progression of sub topics in the document, and users may easily select sub topics by selecting the icons. For example, the overall content of a document

25

might be displayed for each file in a Windows Explorer listing of files. "Overall content" might be determined by examining all the words in the entire document.

Progressive content, represented by several icons, can be displayed next to paragraphs displayed in a display program (e. g. word processor, browser, etc.) or as a sequence of icons displayed elsewhere on the user's graphical user interface.

30

This method provides a visual mechanism for locating and understanding the idea of a document, the location of files, in a user hard disk and their content.

This method can also help many people because very many computer users often
5 make simple mistakes that can take from a few minutes or to an hour to fix. By
making this easy method of choosing a desired file without the need to check to see if
it's the file needed (and then only to discover that it's the wrong file and then having
to go searching for the right file) the method is a very fast and effective way to operate
a user's desktop.

10

The way this method of easy access to files is used is by first having all the
information about a file summed up and then put into a simple phrase that includes
what the content of the file is. The icon of the file is also somehow relevant to what it
contains. This is done so the user can just glance at a file and be able to define what
15 the file contains and if it is what the user is looking for.

Further benefits and advantages of the invention will become apparent from a
consideration of the following detailed description, given with reference to the
accompanying drawings, which specify and show preferred embodiments of the
20 invention.

Brief Description Of The Drawings

Figure 1 is a block diagram illustrating the present invention.

25

Figure 2 describes the structure of a semantic content extractor that may be used in the
practice of this invention.

Figure 3 illustrates the structure of an icon creator that may be used in this invention.

30

Figure 4 shows how a person with a reading disability can use the icon system of this invention.

Figure 5 gives an example of composite icons that represent multiple topics.

5

Figure 6 is a flow chart illustrating a method embodying this invention.

Detailed Description Of The Preferred Embodiments

- 10 Figure 1 is a block diagram explaining the icon process. 108 is a computer that represents a group of directories. 100 represents one directory in one location, 101 represents a second directory in a second location, and 102 represents a third directory in a third location. Each directory has a group of files listed as file 1, file 2, and so on. In a computer, a module is running, 103, the Semantic Content Extractor. 103 can
15 exist within a user's computer, but in this drawing it exists in a server connected to a network, 109. 103, has a running CPU, which extracts the information and content from all the files, 100-103. 104 is responsible for creating an icon using the information provided by the Semantic Content Extractor. Icons may include advertisements, for example if the content is IBM computers, an ad for IBM
20 computers may be presented, and the ad may be a hyperlink to IBM's WEB page. These icons can also be on a separate server as the Semantic Content Extractor. In order to create icons, 104, uses the database of icons, 106, which has a thorough list of icons. The database of icons, 106, is connected to the network, 109. The icons are created by creator of icons module 104. In the module 105, the index of icons to files
25 (or parts of icons to different parts of a text) is created. This indexer module 105 can also be located on a server. The indexer creates an icon and attaches it to a file, 110 and 111.

- 30 There are numerous methods for extracting the content or topic of text documents or portions of documents. These methods include identifying the number of times a

particular word appears in a text or by latent semantic indexing as is known to those skilled in the art.

- Figure 2 describes the structure of the Semantic Content Extractor. This is responsible
- 5 for being able to choose appropriate data to be able to make an icon. 200 represent the input text in a file. 210 determines the size of the text. This can be done by checking the byte size of the file. 201 counts words and characters that can be added up to create an approximation for byte size. In order to speed up the counting process, a key word counter 202 can be created to count key words. Key words are words that are
- 10 essential to represent the meaning of the file. Key words do not include words that are typical for any file (such as and, or, but, the, and so on.). 207 is a database of key words that is created from other documents. 203 speeds up the keyword counting process by counting key phrases used in the text. 205 represents the database of key phrases which holds all key phrases that were obtained from a training database (or
- 15 from processing textual files in past). 204 produces LM from counts that were produced by counting modules 210, 201, 202. The process of making language models (LMs) from counts is described in the reference: Frederick Jelinek, "Statistical Methods for Speech Recognition", The MIT Press, Cambridge, Massachusetts 1998.
- 20 206 is the database of language models that were created from different texts (belonging to different topics - for each topic one LM is made. For example, LM on a medical topic, LM related to travels etc.). 220, is the topic identifier. It defines which language models provide higher likelihood scores (or likelihood ratios) for input texts 200. Since each LM is associated with a topic, it allows to classify each
- 25 textual part with the topic.

If there are several parts in the text that are marked with the different topics, it can be used to associate several topics with the text and make a composite icon that points to different parts of the text with different topics.

The method for classification and segmentation of a text by topics using likelihood ration is described in the patent application no. 09/124,075, for "Real Time Detection of Topical Changes and Topic Identification via Likelihood Based Methods", filed on
5 July 29, 1998.

This process will help create a composite icon, which will allow a better access. 208, the file topic divider, divides the files into their necessary parts and helps create an icon. 209 creates an index of icons to files or an index of parts of icons to different
10 parts of the text.

Figure 3 illustrates the structure of the icon creator. 300 contains topics that were within the Semantic Content Extractor. Topics 1 through 3 have weights listed under them. These weights stand for the importance and significance of topics that are associated with a file. 301 is the intelligent matcher that creates a match of data and images to create an icon. This is done using the database of images 303 and the database of icons 304. The database of images is used only if there is no matching icon for the data given. For example, if there were a topic concerning a car, the computer would search through the database of icons 304. If an icon were not found,
15 one would be created using the database of images 303. 302 extracts an icon that best fits the data given and then creates it to fit a desktop or directory. 306, according to the weight 305 of a topic, the icon combiner creates similar topic icons according to their weight and content. 307, each icon has an index attachment. This attachment to the file opens directly to the file, thus creating easy access to any desired file. This
20 method for opening a file is very effective. Although for blind people, another method of opening files can be created. A blind person can use a sound icon using the database of sound icons 308. This would enable the blind user to use their sense of hearing to choose the files they wish to open.
25

Figure 4 gives an illustration of how a person with a reading disability can use this icon system. 400 is a group of files that are formed into an icon attachment 401. The user then chooses an icon 402, using the pictures or sounds, and the user can then use a speech synthesizer 403 and can listen to a file.

5

Figure 5 gives an example of composite icons that contain multiple topics. 500 shows an icon containing multiple topics, such as cars and travel, 501 and 502, and dealerships 503. 501, the larger part of the file shows cars, the smaller part of the file shows travel. The intermediate sized part of the file shows dealerships. 503 contains 10 an index which lists information on cars or buildings 506. 504 shows where the information on cars is placed in the file. Using a fraction method, the files can be broken down, as shown in 504 and 505. 510 shows the file.

Figure 6 shows a flowchart of the method. At 600, a list of files is generated. Step 15 601 reads the content of each file, and at 602, topics are attached to each file. At 603, icons are generated for files. At 604, if several topics, a composite icon is created containing many topics. At 607, an index of topics is created. At 605, a list of icons is printed near file names. At 606, a list of icons can be created to list files.

20 While it is apparent that the invention herein disclosed is well calculated to fulfill the objects stated above, it will be appreciated that numerous modifications and embodiments may be devised by those skilled in the art, and it is intended that the appended claims cover all such modifications and embodiments as fall within the true spirit and scope of the present invention.

25

25